

Mushroom classification EDA VSCode Notebook

```
In [ ]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
plt.style.use("ggplot")
```

```
In [ ]: data_location = "../data/mushrooms.csv"
mushrooms = pd.read_csv(data_location)
mushrooms.head(5)
```

Out []:

	class	cap- shape	cap- surface	cap- color	bruises	odor	gill- attachment	gill- spacing	gill- size	gill- color	...	stalk- surface- below- ring	stalk- color- above- ring	stalk- color- below- ring	veil- type	veil- color	ring- number	ring typ
0	p	x	s	n	t	p	f	c	n	k	...	s	w	w	p	w	o	
1	e	x	s	y	t	a	f	c	b	k	...	s	w	w	p	w	o	
2	e	b	s	w	t	l	f	c	b	n	...	s	w	w	p	w	o	
3	p	x	y	w	t	p	f	c	n	n	...	s	w	w	p	w	o	
4	e	x	s	g	f	n	f	w	b	k	...	s	w	w	p	w	o	

5 rows × 23 columns



```
In [ ]: mushrooms.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8124 entries, 0 to 8123
Data columns (total 23 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   class                                8124 non-null   object
1   cap-shape                            8124 non-null   object
2   cap-surface                          8124 non-null   object
3   cap-color                           8124 non-null   object
4   bruises                             8124 non-null   object
5   odor                                8124 non-null   object
6   gill-attachment                      8124 non-null   object
7   gill-spacing                        8124 non-null   object
8   gill-size                           8124 non-null   object
9   gill-color                          8124 non-null   object
10  stalk-shape                         8124 non-null   object
11  stalk-root                          8124 non-null   object
12  stalk-surface-above-ring            8124 non-null   object
13  stalk-surface-below-ring            8124 non-null   object
14  stalk-color-above-ring              8124 non-null   object
15  stalk-color-below-ring              8124 non-null   object
16  veil-type                           8124 non-null   object
17  veil-color                          8124 non-null   object
18  ring-number                         8124 non-null   object
19  ring-type                           8124 non-null   object
20  spore-print-color                   8124 non-null   object
21  population                          8124 non-null   object
22  habitat                             8124 non-null   object
dtypes: object(23)
memory usage: 1.4+ MB
```

No null values, all features are categorical.

```
In [ ]: # mappings written manually based on Kaggle data explorer
feature_value_mappings = {
    "class": {"e": "edible", "p": "poisonous"},
    "cap-shape": {"b": "bell", "c": "conical", "x": "convex", "f": "flat", "k": "knobbed", "s": "sunken"},
    "cap-surface": {"f": "fibrous", "g": "grooves", "y": "scaly", "s": "smooth"},
    "cap-color": {"n": "brown", "b": "buff", "c": "cinnamon", "g": "gray", "r": "green", "p": "pink", "u": "purple", "e": "red"},
    "bruises": {"t": "bruises", "f": "no"},

```

```

"odor": {"a": "almond", "l": "anise", "c": "creosote", "y": "fishy", "f": "foul", "m": "musty", "n": "none", "p": "pungent"},
"gill-attachment": {"a": "attached", "d": "descending", "f": "free", "n": "notched"},
"gill-spacing": {"c": "close", "w": "crowded", "d": "distant"},
"gill-size": {"b": "broad", "n": "narrow"},
"gill-color": {"k": "black", "n": "brown", "b": "buff", "h": "chocolate", "g": "gray", "r": "green", "o": "orange", "p": "pink", "e": "red"},
"stalk-shape": {"e": "enlarging", "t": "tapering"},
"stalk-root": {"b": "bulbous", "c": "club", "u": "cup", "e": "equal", "z": "rhizomorphs", "r": "rooted", "?": "missing"},
"stalk-surface-above-ring": {"f": "fibrous", "y": "scaly", "k": "silky", "s": "smooth"},
"stalk-surface-below-ring": {"f": "fibrous", "y": "scaly", "k": "silky", "s": "smooth"},
"stalk-color-above-ring": {"n": "brown", "b": "buff", "c": "cinnamon", "g": "gray", "o": "orange", "p": "pink", "e": "red"},
"stalk-color-below-ring": {"n": "brown", "b": "buff", "c": "cinnamon", "g": "gray", "o": "orange", "p": "pink", "e": "red"},
"veil-type": {"p": "partial", "u": "universal"},
"veil-color": {"n": "brown", "o": "orange", "w": "white", "y": "yellow"},
"ring-number": {"n": "none", "o": "one", "t": "two"},
"ring-type": {"c": "cobwebby", "e": "evanescent", "f": "flaring", "l": "large", "n": "none", "p": "pendant", "s": "sheathing"},
"spore-print-color": {"k": "black", "n": "brown", "b": "buff", "h": "chocolate", "r": "green", "o": "orange", "u": "purple"},
"population": {"a": "abundant", "c": "clustered", "n": "numerous", "s": "scattered", "v": "several", "y": "solitary"},
"habitat": {"g": "grasses", "l": "leaves", "m": "meadows", "p": "paths", "u": "urban", "w": "waste", "d": "woods"}
}
for column, mapping in feature_value_mappings.items():
    if column in mushrooms.columns:
        mushrooms[column] = mushrooms[column].replace(mapping)
    else:
        print(f"Column {column} not found in mushrooms.csv")
mushrooms.head(5)

```

Out[]:

	class	cap- shape	cap- surface	cap- color	bruises	odor		gill- attachment	gill- spacing	gill- size	gill- color	...	stalk- surface- below- ring	stalk- color- above- ring	stalk- color- below- ring	veil- type	veil- color
0	poisonous	convex	smooth	brown	bruises	pungent		free	close	narrow	black	...	smooth	white	white	partial	white
1	edible	convex	smooth	yellow	bruises	almond		free	close	broad	black	...	smooth	white	white	partial	white
2	edible	bell	smooth	white	bruises	anise		free	close	broad	brown	...	smooth	white	white	partial	white
3	poisonous	convex	scaly	white	bruises	pungent		free	close	narrow	brown	...	smooth	white	white	partial	white
4	edible	convex	smooth	gray	no	none		free	crowded	broad	black	...	smooth	white	white	partial	white

5 rows × 23 columns



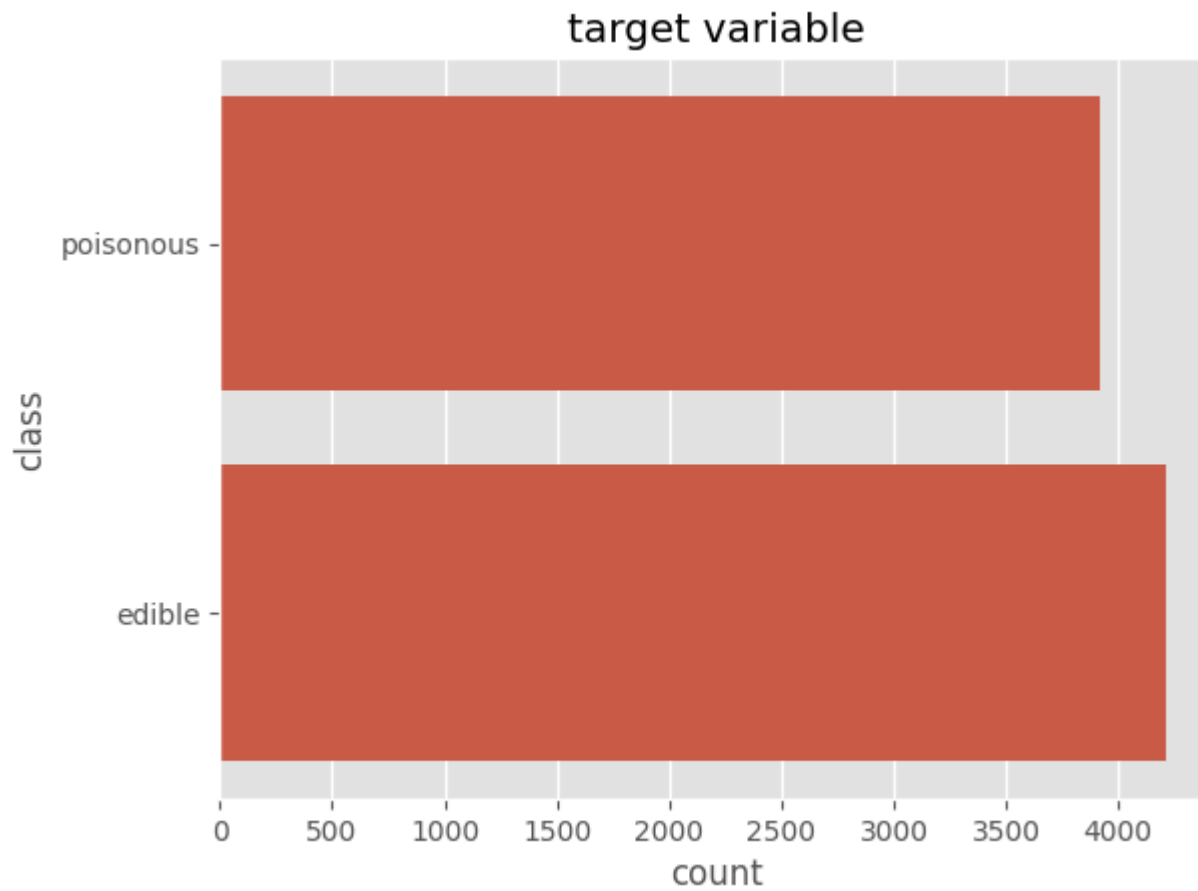
In []:

```
print(mushrooms["class"].value_counts())
print()
print(mushrooms["class"].value_counts() / len(mushrooms))
sns.countplot(mushrooms["class"])
plt.title("target variable")
```

```
class
edible      4208
poisonous   3916
Name: count, dtype: int64
```

```
class
edible      0.517971
poisonous   0.482029
Name: count, dtype: float64
```

Out[]: Text(0.5, 1.0, 'target variable')



Target variable analysis

Fairly balanced target variable (51% edible 48% poisonous).